

⑤ Int. Cl.²
G 10 L 1/04

⑥ 日本分類
96(3) A 02

⑦ 日本国特許庁

⑧ 特許出願公告

昭50—1846

特 許 公 報

⑨ 公告 昭和50年(1975) 1月22日

発明の数 1

(全 10 頁)

1

④ 音声認識装置

⑪ 特 願 昭 4 5 - 4 1 6 4 9
⑫ 出 願 昭 4 5 (1 9 7 0) 5 月 1 8 日
⑬ 発 明 者 中田和男
国分寺市東恋ヶ窪 1 の 2 8 0 株
会社日立製作所中央研究所内
同 市川煮
同所
同 藤村靖
鎌倉市長谷 6 2 2
⑭ 出 願 人 株式会社日立製作所
東京都千代田区丸の内 1 の 5 の 1
⑮ 代 理 人 弁理士 薄田利幸

図面の簡単な説明

第 1 図は本発明に用いる声帯振動検出器の取付位置を示す図、第 2 図は上記検出器によつて検出された声帯振動波形の一例を示す図、第 3 図は本発明に用いる鼻音出力検出器の取付位置を示す図、第 4 図は上記鼻音出力検出器によつて検出された波形図の一例を示す図、第 5 図は本発明に用いられる人工口蓋と接触位置検出用電極配置図、第 6 図は上記第 5 図装置による出力特性図、第 7 図は本発明に用いる唇閉鎖検出器の取付を示す図、第 8 図は上記第 7 図に示す検出器による出力特性の一例を示す図、第 9 図は本発明音声認識装置の一実施例の構成を示すブロック図である。

発明の詳細な説明

本発明は音声の認識装置、とくに日本語音声の認識に係り、話者の調音運動を非音響的に表わす情報を利用する音声認識装置に関するものである。

従来知られている音声認識装置では、そのほとんどが分析される音声の情報源を、マイクロホンによつて抽出された音響情報に求めている。そのため、(1)ある種の音素(または音節)について識別が難しい。たとえば、鼻音の検出と〔m〕、

2

〔n〕、〔o〕の相互の識別、破列音〔p〕、〔t〕、〔k〕、〔b〕、〔g〕の検出と相互識別など非常に難しい。(2)装置の使用環境によつてはマイクロホンによつて得られる情報のみでは信号対雑音比が低く、エネルギー強度の低い子音の情報を正確に抽出できず、正確な認識、識別ができない。

本発明は上述のような従来装置の欠陥を克服するとともに、正確に音声の識別が行える音声認識装置を提供することを目的とするものである。以下本発明について説明する。

本発明は音声を識別するために、その音声の発声音の声帯振動、鼻孔出力の有無、唇の完全閉鎖の有無および舌と口蓋の接触による舌の位置の時間的变化の 4 つの情報を直接発声者から抽出し、別のマイクロホン等によつて抽出された音響的な音声出力からの情報と組合せることによつて正確な音声認識装置を実現するものである。

音声の発声は本来生理的現象であつて、音声波は発声器官の調節運動の結果生じた物理的な音響振動である。したがつて、音声学によつて体系づけられているように、音声波の特徴はその調音の仕方に起因しており、その調音機構の情報を直接、測定、抽出することによつて、音響的な音声波出力の情報のみからは得がたい子音についての特徴を抽出することができる。以下、最初に各情報の検出器の機能について説明する。

- (1) 声帯振動情報の抽出、第 1 図は声帯振動検出器である感圧素子(たとえば感圧トランジスタ)の取付を示すもので、図示のように、感圧素子 2 を気管の前壁、声帯 1 の近くに圧着して設定されており、発話の内容と無関係に声帯の振動を検出することができる。第 2 図は上記検出器 2 の出力をそのときの音響波形と対比して示したもので、数の 8〔hat / i〕と 4〔j o N〕の発音を例として示したものである。
- (2) 鼻孔出力情報の抽出、第 3 図は鼻孔出力検

3

出器である感圧素子3の取付を示すもので、図示のように感圧素子3を鼻壁の鼻孔近くに圧着して取付けられており、発話の内容と無関係に鼻孔出力の有無、すなわちその音が鼻音(鼻音化されたもの)であるか否かを検出することができ

る。第4図は上記鼻孔出力検出器3の出力を、その検出時の音響波形と対比して示したもので、数の3[s a N]と7[n a n a]の例について示したものである。

(3) 舌と口蓋の接触状態による調音とその時間的変化、とくに舌による口腔内の完全閉鎖あるいは狭窄の有無とその位置の情報の抽出、第5図は上記舌と口蓋の接触位置を検出する装置ダイナミック・パラトグラフ(dynamic palatograph 例えば文献S. Shibata, "A study of Dynamic Palatography" Ann. Bull. No.2 Res. Inst. Logopediss and phoniatics 第28~36頁等がある)の人工口蓋の電極の配列を示す図で、この装置によつて、発音時における舌と口蓋の接触状態による舌の調音、とくに舌による口腔内の完全閉鎖あるいは狭窄の有無とその位置ならびにその時間的変化を抽出する。第6図は上記パラトグラムによる検出出力の波形を示すもので、歯茎部で完全閉鎖が起っている例[a d a]、[a t a]、[a t s a]、および狭窄(せばめ)はあるが完全閉鎖になつていない例[a s a]、[a f a]について示している。

(4) 唇の調音とくにその完全閉鎖の有無の情報の抽出、第7図は上記完全閉鎖の有無の情報を抽出するための検出器の取付を示すもので、図示のように、上下両唇に相対して小さな電極をおき、その接触によつて唇の完全閉鎖の有無の情報を得るようにしたものである。第8図は上記唇の閉鎖情報検出器による検出信号を、その音波と対比して示したもので、語ナマ(nama)、サバク(s a b a k w)について示したものである。

本発明は上記のように発声による声帯振動の検知、鼻音化による鼻音出力の検知、調音運動にもとづく情報(調音運動の状態およびその時間的な変化の当該器官の接触の検知によつてえられる情

4

報)に必要な応じて音響信号すなわちマイクロホンからの音声波の情報を抽出し(その音波を構成する適当な周波数成分の検出)その目的によりこれらの情報の適当な組合せにより、音声識別(認識)を行なうものである。

第9図は本発明音声認識装置の一実施例の構成を示すブロック図である。

音声波による音響情報は接話マイクロホン4によつてできるだけ室内の騒音をさけ、しかも、高周波のエネルギーを失なうことなく抽出され、前置増幅回路7で適当なレベルに増幅された後、包絡波(エンベロープ)検出回路12、ホルマント周波数抽出回路13、雑音成分抽出回路14に分配される。

一方、声帯振動の情報が気管前壁波マイクロホン2によつて検出され、前置増幅器9で増幅された後、整形回路15によつて適当な閾値で整形され、声帯振動の有無の信号として音素分類回路19に送られる。

鼻音の抽出のため鼻孔出力の有無が鼻壁に固定した検出器16によつて鼻孔出力の有無の信号として音素分類回路19に加えられる。

また、唇の完全閉鎖を示す情報が両唇に取付られた検出器5によつて検出され、前置増幅器10、閾値整形回路17によつて、唇の完全閉鎖の有無の信号として音素分類回路19に加えられる。

最後に、舌と口蓋の接触状態とその時間的変化による舌の調音の情報、とくに舌による口腔内の完全閉鎖の有無およびその位置が歯茎部であるか口蓋部であるかを示す情報がパラトグラム測定装置6によつて検出され、前置増幅器11で増幅された後、情報解読回路18によつて選ばれた歯茎口蓋部位における舌の接触の有無、その時間的変化を表す信号として音素分類回路19に加えられる。

上記音素分類回路では上記各入力2値信号、すなわち、各サンプル時点における信号の有無の状態によつて論理的に音素を分類してゆく。上述の5種類の測定結果を要因として日本語音素の分類を行うとその音声区分の定常部時点での判定基準は第1表のようになる。

第 1 表

アイウエオ 表記による 分類	対応する 音素記号	気管前 壁波に よる有 声音	雑音成 分によ る無声 音	鼻壁振 動によ る鼻音	舌の接触点 a b c	リップ 中央部 接触
母音(イを 除く)	(v)	+	-	-	- - -	-
ヤ行 子音 及び イ	j	+	-	-	- + -	-
ワ行 子音	w	+	-	-	- - -	-
ラ行 "	ℓ r	+	-	-	+ - -	-
マ行 "	m	+	-	+	- - -	+
ナ行 "	n	+	-	+	+ - -	-
ン	N	+	-	+	± - ±	-
バ行 子音	p	-	-	-	- - -	+
タ行 "	t ts	-	-	-	+ - -	-
カ行 "	k	-	-	-	- - +	-
パ行 "	b	±	-	-	- - -	+
ダ行 "	d	±	-	-	+ - -	-
ガ行 "	g	±	-	±	- - +	-
サ行 "	s f	-	+	-	- + -	-
ザ行 "	z B	±	-	-	- + -	-
ハ行 "	∅	-	+	-	- - -	-

上記表において「+」は接触検出、「±」は検出不確実、「-」は接触なしを表わす。
 上記表より、準定常的または比較的ゆるやかな過渡音の母音、半母音(i, n)、流音の群(ℓ, r)と雑音部のエネルギーが強く音響的な特徴の抽出が確実にできる無声まっつ音(s, f等)の第1群と、それ以外の子音の第2グループに大別できる。この第2群の子音が従来の純音響的情報からのみ特徴を抽出するものでは最も認識、識別が困難とされていた。

* 上記分類表に示される各音素は、さらに第2表に示すような分類論理によつて分類でき、上記第2群に属する音素は撥音N(日本語「ン」に対応)を除いては完全に分離識別できる。
 撥音Nは前後の環境によつてm, n, Bの鼻音のいずれかまたは単に母音の鼻音化として調音される。したがつて第2表に示すように分類されるが、その結果から前後関係単語としての言語情報によつてNと、m, n, Bは容易に分離できる。

第 2 表

(第1群)

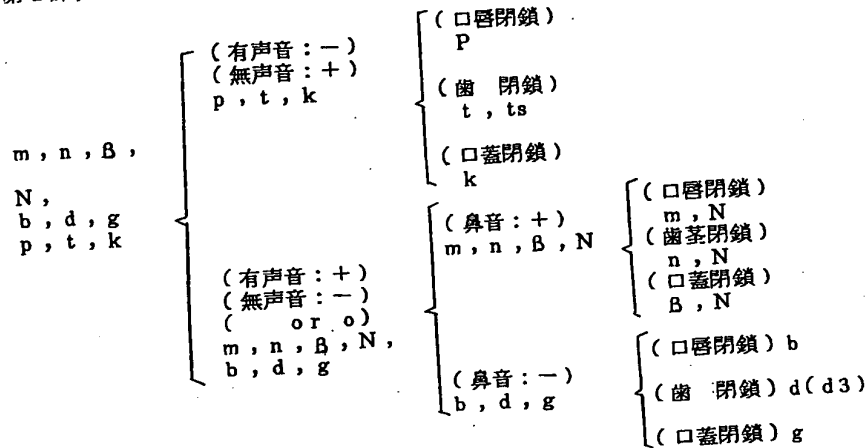
v, j, w, ℓ, r, N, s, f, z, B, ∅

(有声音: +)
v, j, w, ℓ, r, N, z, B
(無声音: +)
s, f, ∅

(鼻音: +)
N
(鼻音: -)
v, j, w, ℓ, r, z, B

(無声音: -)
v, j, w, ℓ, r
(無声音: +)
r, B

〔第2群〕

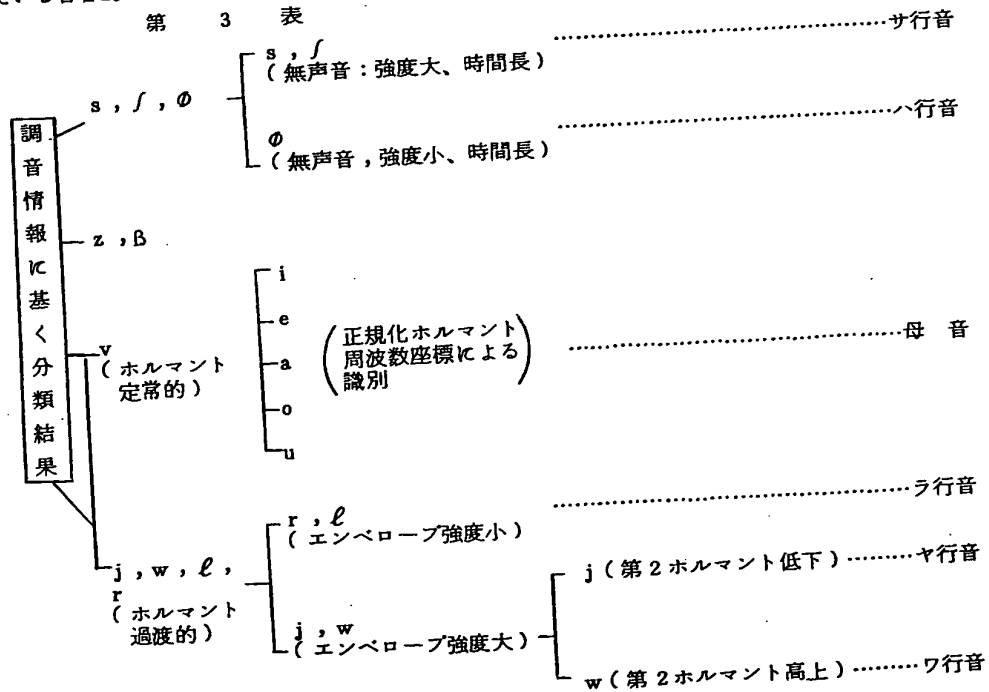


上記表においてBは鼻濁音を表わし、又発音記号の下に棒を引いたものは分類された最終的出力を表わす。

上記第1群のさらに細かい分類は、従来知られていない音響的な識別原理に基いて、ホルマント周 * 20

* 波数とその時間的な変化、雑音部分の強度、時間長等の要因によつて分類識別する。第3表はこれらの、音響特性による日本語音素の分類を示すものである。

第3表



9

なおこの中で r または l は音響的に比較的に識別し難い音素であるが、上記パラトグラフによる舌の調音情報とくにその時間的变化パターンを併用する事によつて検出が確実になる。また、第6図の に示す音声 [a t a] と [a t s a] のパラトグラフの比較からみても分るように、はれつ音とはさつ音の区別にもパラトグラフによる情報を役立てることができる。

上述する分類論理によつて、各特徴抽出器の出力より、分析される音素が第1群でかつ鼻音でないことが決定されると、その時点で音響音素分類回路20が駆動され、その分類回路20は上記第3表に示した論理によつて分類し、その結果は端子22-1ないし22-10に出力される。すなわち、それぞれの端子の出力は音素 [s | f] , [φ] , [i] , [e] , [a] , [o] , [w] , [r | l] , [j] , [w] に対応する。なお i ではよう音 [j] + 母音として識別される。

又分析される音素が第2群に属することが判明すると、音素分類回路19が駆動され、上記第2表に示した分類論理によつて音素の分類を行う。第2群に属する音素は日本語の音韻表記としては第2表にした分類で十分な細さで分類識別され、音素分類回路19の出力端子21-1ないし21-9に出力される。ここで端子21-1ないし21-9の出力は、それぞれ [p] , [l] , [k] , [m | N] , [n | N] , [o | N] , [b] , [d] , [g] に対応する。

上記装置の構成において音素間の区分は、発語の原因となる調音の情報を直接検出しているもので、

10

これらの情報の検出結果 (" 1 " , " 0 " 信号) の変化点で区分が行なわれることになる。このことは音響的な情報のみから音声識別を行なつていた従来装置にくらべ、はるかに容易に区分することができ。

上記実施例の説明では電話等の音声情報のみによる音声源からの音声識別は不可能であるが、音声と共に、他の特徴抽出情報 (声帯振動、鼻外壁振動等) を送信側で抽出し、音声帯域内の副搬送波にのせて伝送できるから、遠隔地の話者の音声認識にも適用できるものである。

又、本発明音声認識装置において、識別結果に単語としての言語情報からの推定を加え、その不確実さがある場合の誤りを訂正し得ることは従来の音声認識装置と同様に実現できる。

⑨特許請求の範囲

1 声帯振動の有無検出器と鼻孔出力の有無検出器と唇の閉鎖の有無検出器と舌と口蓋の接触の有無検出器の各出力を有無を表す2値信号に変換する回路と、音響信号抽出マイクロホンからえられる音声信号から雑音信号の有無を検出し2値信号に変換する回路と、上記各2値信号を組合せて特定子音を識別する第1の音素分類回路と、上記マイクロホンから得られる音声信号の包絡波、ホルマント周波数、雑音成分を抽出し、その各出力と上記音素分類回路出力の1部と組合せ上記第1の音素分類回路以外の音素を識別する第2の音素分類回路とを含んで構成されることを特徴とする音声認識装置。

図 1

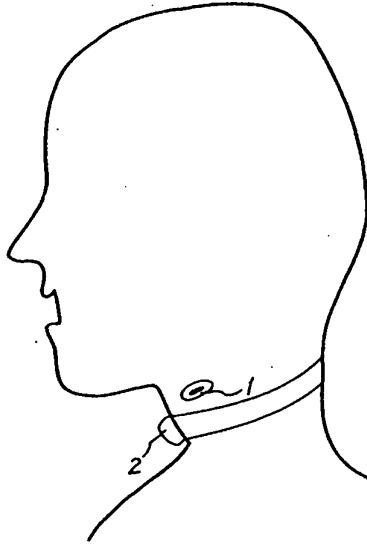


図 2

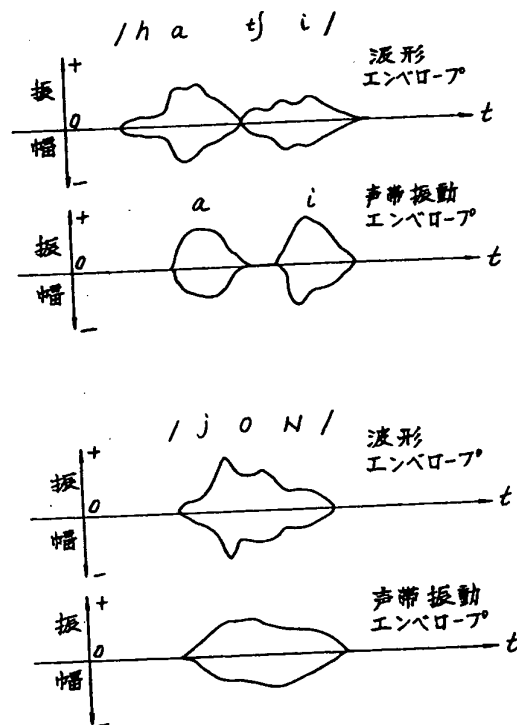


図 3

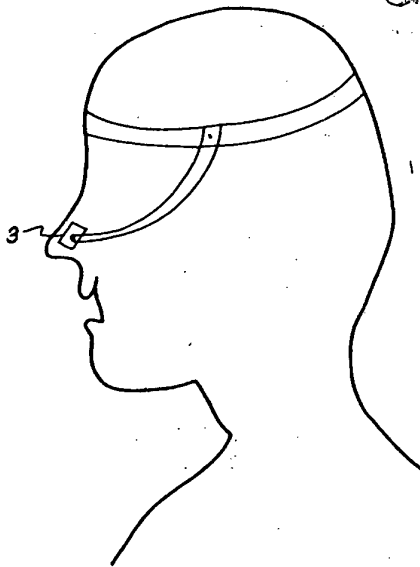


図 4

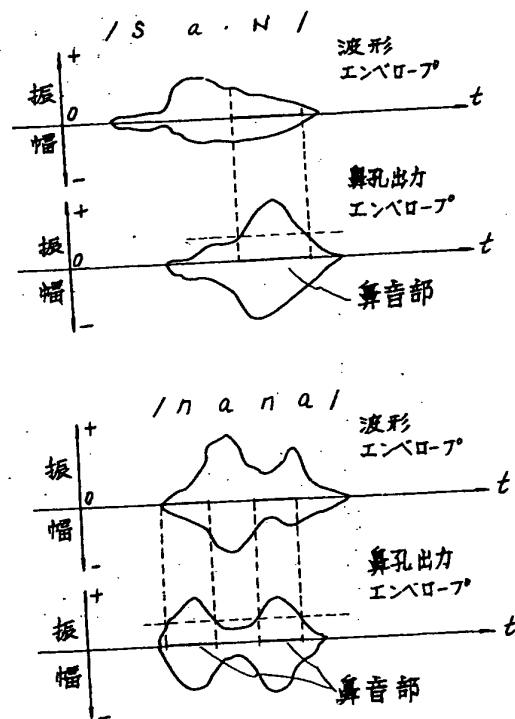


図 5

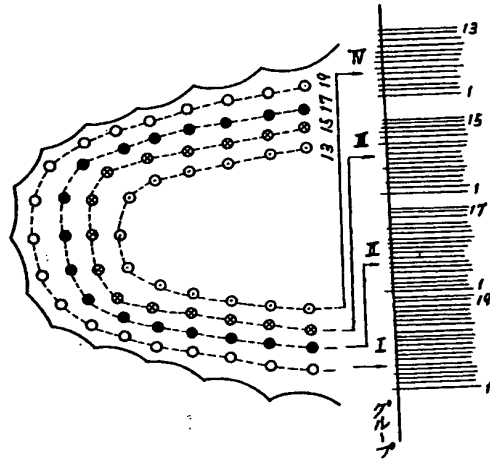
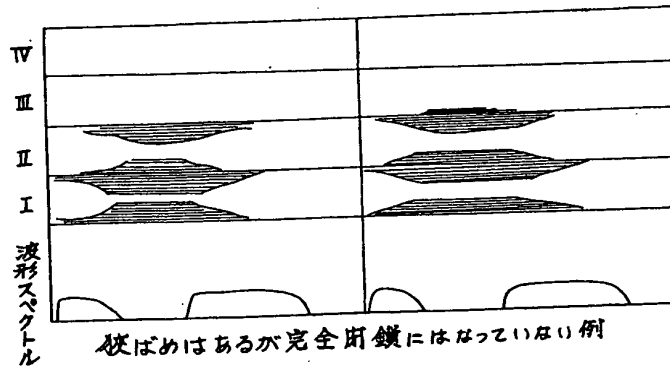
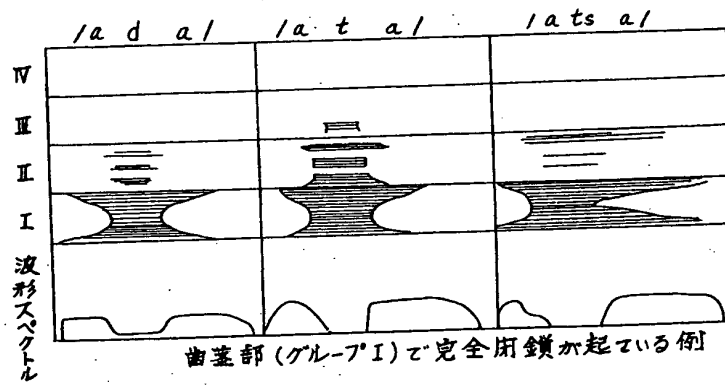
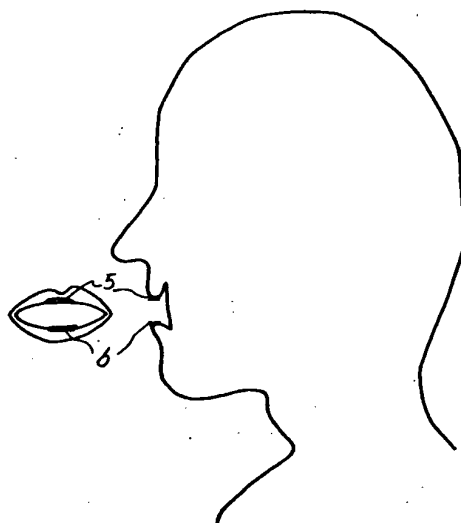


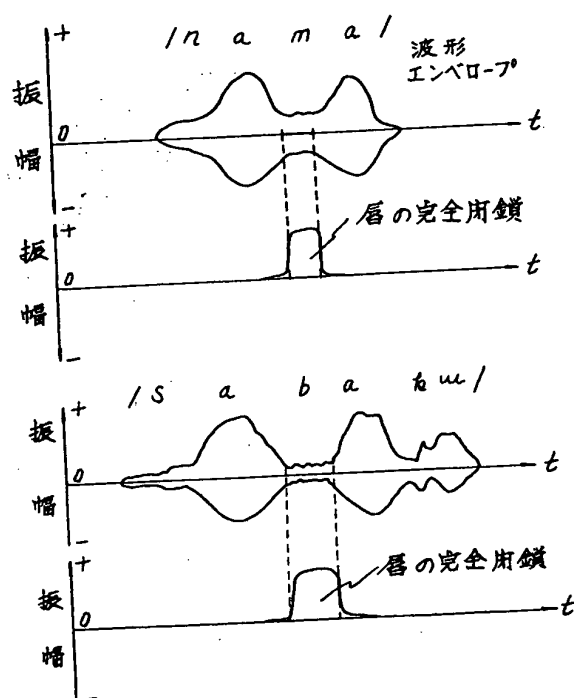
図 6



和 7 図



和 8 図



Abridgement of Reference (1)

- (1) Japanese Patent Examined Publication No. 50(1975)-1846
- (2) Date of Examined Publication: January 22, 1975
- (3) Application No.: 45(1970)-41649
- (4) Filing Date: May 18, 1970
- (5) Inventors: Kazuo Nakada et al.
- (6) Applicant: Hitachi Seisaksho Co., Ltd. (Japan)

Title of Invention:

"Fonometer"

Brief Explanation of Drawings:

Figs. 1 to 9 show an embodiment of the invention.

Abridgement

1 --- vocal cord 2 --- pressure sensing element
for detecting vibration of the vocal cord 1 with no relation
with the contents of the conversation issued 3 --- pressure
sensing means sector for detecting output from nares
4 --- microphone for picking up acoustic information in voice
wave 5 --- detector for detecting complete closure of lips
6 --- palatogram measuring device 7, 10 --- preparatory
amplifying circuit 12 --- envelope detecting circuit
14 --- noise component detecting circuit 15 --- shaping
circuit 16 --- detector for detecting output from nares
17 --- threshold shaping circuit 18 --- information decoding
circuit 19 --- sound element analyzing circuit

According to the invention, in order to identify or distinguish voice, four information, that is, vocal cord vibration of vocalized (or issued) sound of the voice, existence or non-existence of output from nares, existence or non-existence of complete closure of lips and output

from nares, temporal variation of position of lingua due to the contact between the lingua and palate are extracted directly from a person who is vocalizing, and said information is combined with another information from acoustic voice wave output which is extracted from another microphone etc., to thereby obtain an accurate voice identifying apparatus (i.e., fonometer).